

Personalización y Sincronización de Subtítulos en Escenarios Multi-Pantalla

Zuzanna Parcheta¹, Mario Montagud^{1,2}, Jordi Belda¹, Fernando Boronat¹

¹ Immersive Interactive Media (IIM) R&D Group

¹ Campus de Gandia - Universitat Politècnica de València (UPV), Spain

² Centrum Wiskunde & Informatica (CWI), Amsterdam (The Netherlands)

{zupar@epsg., mamontor@, jorbelva@epsg., fboronat@dcom.}upv.es

Keywords: Accessibility, IDES, Multi-Screen, Social Integration, Subtitles, Synchronization, Web

Abstract. This paper presents a web-based platform that enables the customization and synchronization of subtitles when using both a shared main screen and multiple secondary screens. This platform enables the dynamic adaptation of the subtitles' format (font family, size, color...) and positioning in order to maximize the users' perceived Quality of Experience or QoE (e.g., according to their preferences and/or needs). Likewise, the users can dynamically select the subtitles' language, apply a customized delay offset to the subtitles and adjust the number of lines to be displayed. Moreover, by using a secondary device, each user can also consume the subtitles in a customizable manner, by additionally enabling their presentation in two different languages and by navigating between them, being able to restore the video playback position by clicking on a specific subtitle line. With the provided functionalities, our platform enables personalized and immersive media experiences, and can contribute to a better language learning and social integration of audience with audiovisual impairments and non-natives, in both domestic and crowded multi-culture environments (e.g., public transportations systems, touristic places...).

Introduction

Los contenidos e interfaces de los servicios multimedia deben ser adaptativos y personalizables con tal de proporcionar información útil, amigable y comprensible para todos los usuarios. En este contexto, los subtítulos y otros tipos de datos multimedia textuales, como descripciones o metadatos, son componentes clave para proporcionar servicios multimedia enriquecidos, personalizados y accesibles. Los subtítulos constituyen una vía rápida y eficiente de acceso e interpretación del contenido multimedia para un porcentaje significativo de la audiencia televisiva y de contenidos audiovisuales en general. Es por ello que las instituciones internacionales y los proveedores de contenidos se están esforzando para incrementar el porcentaje de servicios multimedia subtitulados. Como prueba de evidencia, la cadena BBC intenta subtitular el 100% de sus contenidos [1].

La aplicabilidad y beneficios aportados por los subtítulos son múltiples. En primer lugar, los subtítulos son muy útiles para los usuarios con deficiencias audiovisuales. Por un lado, los usuarios con deficiencias auditivas pueden acceder (mejor) a la información del audio a través de los subtítulos. Por otro lado, la personalización del formato de los subtítulos (p.ej., tamaño, tipo de fuente, color...) puede ser muy útil para personas con deficiencias visuales. Sin embargo, los subtítulos no son únicamente beneficiosos para romper barreras audiovisuales, sino que su aplicabilidad se extiende a otros ámbitos de la integración social, ya que son herramientas muy útiles para los consumidores que no entiendan el idioma (o el acento) del audio o que tienen dificultades de comprensión [2, 3]. Además, los subtítulos son muy útiles en otros entornos de aprendizaje, terapia del lenguaje y alfabetización [4],

así como también constituyen herramientas de apoyo para estimular el procesamiento cognitivo y para captar la atención de los consumidores [5]. Finalmente, aparte de las ventajas anteriores, varios estudios (p.ej., [2, 3, 6-9]) han demostrado que una presentación dinámica y personalizable de los subtítulos es muy conveniente para mejorar el confort, la inmersividad (*engagement*) y, en general, la calidad de experiencia (QoE, *Quality of Experience*) percibida por los usuarios cuando consumen contenidos audiovisuales.

Aunque los subtítulos se consuman principalmente en entornos personales y domésticos, también son muy beneficiosos en sitios en los que coincide gente que habla diferentes idiomas (p.ej., sitios turísticos, estaciones o sistemas de transporte público...) o sitios muy concurridos en los que no se pueda escuchar el audio o no sea conveniente un volumen alto (p.ej., restaurantes, museos, hospitales...).

Debido a los múltiples beneficios y la amplia aplicabilidad de los subtítulos, un porcentaje significativo de usuarios los consume, pero este porcentaje sería mucho mayor si los subtítulos estuviesen disponibles en más servicios [7]. Además, las plataformas de subtítulos existentes no explotan al máximo las posibilidades de adaptabilidad y accesibilidad. Como respuesta a estas carencias, este artículo presenta una plataforma web que proporciona sincronización y personalización de subtítulos en entornos multi-pantalla. Una característica clave de esta plataforma es que se ha implementado mediante el uso exclusivo de tecnologías web estándar, como HTML5 y Javascript. Esto garantiza soporte multi-red, multi-dispositivo, multi-plataforma y multi-navegador, así como facilita una implantación ubicua. Además, únicamente se necesita un navegador web para utilizar esta plataforma, sin requerir la instalación de ningún software o hardware adicional. Esta plataforma permite la adaptación y personalización dinámica de la presentación de los subtítulos en base a las necesidades de los usuarios (p.ej., su idioma), capacidades sensoriales (p.ej., adaptando el tamaño, color...) y preferencias (p.ej., idioma, formato, tamaño, posición, número de líneas de subtítulos...), así como también en base a los dispositivos disponibles (p.ej., smartphone, tablet, laptop, TV...), la evolución del contenido o aplicación (p.ej., patrones de conversación, contraste con el fondo de las escenas, re-dimensionado y re-posicionamiento dinámico...) y de las condiciones del entorno específico en los que se consuman, garantizando en todo momento su sincronización con el resto de contenido audiovisual. Además, esta plataforma proporciona otras funcionalidades importantes. En primer lugar, permite aplicar un desfase temporal, positivo o negativo, a la presentación de los subtítulos con respecto a la presentación del vídeo. Esto es útil si los usuarios prefieren leer los subtítulos un poco antes o después que el audio asociado para comprobar cómo se pronuncian ciertas palabras (o frases) o si son capaces de identificarlas, respectivamente. En segundo lugar, se pueden modificar las líneas de subtítulos que se muestran simultáneamente. Esto es útil si los usuarios no son capaces de leer ciertas palabras o frases a la velocidad a la que avanzan los subtítulos. En este caso, la plataforma también permite que los usuarios hagan click sobre una línea de subtítulos específica para saltar a la posición del vídeo correspondiente, por si quieren volver a escuchar una palabra o frase específica. Además, todas las funcionalidades anteriores están disponibles en los dispositivos secundarios, permitiendo que cada usuario consuma de manera personalizada los subtítulos, seleccionando su idioma y configurando el formato más apropiado, de manera sincronizada con el resto de dispositivos en el escenario multi-pantalla. La Sincronización entre Dispositivos (IDES, *Inter-Device Synchronization*) se consigue gracias al diseño de dos componentes: un mecanismo de sincronización de relojes virtual y un protocolo IDES. Asimismo, la plataforma permite compartir los comandos de control de navegación (p.ej., *play*, *pause* y saltos a posiciones específicas) entre todos los dispositivos, manteniendo, en todo momento, la sincronización. Finalmente, en cada dispositivo secundario, los usuarios pueden seleccionar simultáneamente la presentación de los subtítulos en dos idiomas. Esto puede ser muy útil para el aprendizaje de idiomas.

Tal y como se detalla en la siguiente sección, nuestra plataforma presenta múltiples ventajas y funcionalidades añadidas en comparación con otras plataformas de subtítulos existentes. Por tanto, pensamos que puede tener impacto en el paradigma actual de consumo de contenidos multimedia. Aunque su diseño e implementación no estén finalizados, los componentes clave para conseguir las funcionalidades de consumo multimedia, personalización, sincronización e interacción ya están disponibles, y se presentan en este artículo. Además, se pretende mostrar prototipos de la misma en la conferencia. Esto permitirá que la audiencia experimente con ella, así como recibir *feedback* valioso sobre sus aspectos de diseño y funcionalidades futuras.

Estructura del Texto

Este artículo se ha estructurado de la siguiente manera. En la Sección 2 se presentan algunos trabajos relacionados. Los componentes de la plataforma y sus funcionalidades se describen en la Sección 3. Finalmente, en la Sección 4 se presentan algunas líneas de trabajo futuro.

Trabajos Relacionados

En esta sección se presentan algunos estudios que han demostrado los beneficios aportados por la presentación adaptada y dinámica de subtítulos, así como algunas plataformas existentes (para contenido almacenado) relacionadas con la presentada en este artículo, destacando sus ventajas.

En primer lugar, varios estudios (p.ej., [6, 8, 9]) han demostrado que un posicionamiento dinámico de los subtítulos (p.ej., en base al color o fondo de las escenas, o a la posición de los oradores) en comparación con el posicionamiento fijo en la parte inferior de la pantalla es beneficioso para: i) mejorar la accesibilidad para usuarios con deficiencias auditivas; ii) reducir los patrones de movimiento de la mirada, evitando distracciones y reduciendo el cansancio; iii) mejorar la legibilidad; iv) proporcionar una experiencia más inmersiva y emotiva; v) captar la atención de los usuarios de manera más efectiva; vi) proporcionar una solución más estética, contemporánea y creativa; vii) proporcionar mayor naturalidad; y viii) ayudar a los usuarios a asociar mejor los subtítulos con los oradores correspondientes. Además, el estudio en [9] demuestra que los patrones de miradas de los usuarios cuando se utiliza subtítulo dinámico son similares a cuando los subtítulos no están presentes, y proporcionan una mejor QoE para la gran mayoría de usuarios.

Por ejemplo, la plataforma en [6] proporciona subtítulo dinámico mediante técnicas de reconocimiento de rostros, y análisis de movimiento de labios y de prominencia visual. Además, presenta los subtítulos de manera progresiva, palabra por palabra, para un mejor seguimiento de la evolución de la historia y muestra la variación del volumen del audio, lo que proporciona información sobre las emociones de los oradores. Asimismo, la plataforma en [8] presenta dos mejoras relevantes sobre la presentada en [6]. En primer lugar, incorpora un algoritmo más preciso para la detección de oradores, teniendo en cuenta información audiovisual y no solo el movimiento de labios, ya que otras personas en la misma escena podrían también estar moviendo los labios. En segundo lugar, el posicionamiento dinámico de los subtítulos tiene en cuenta la ausencia de oradores en escenas específicas, la distancia entre subtítulos en escenas correlativas, los límites de la pantalla y que no se oculte información importante de las escenas.

La plataforma presentada en [7] permite la personalización y sincronización de subtítulos en entornos multi-pantalla. Esta plataforma presenta los subtítulos, utilizando el formato obsoleto SubRip (.srt), cargándolos en componentes Scalable Vector Graphics (SVG) y la sincronización se consigue tratando estos componentes SVG como una animación Synchronized Multimedia Integration Language (SMIL) adicional. La sincronización IDEs

se consigue compartiendo las líneas de tiempo SMIL entre todos los dispositivos. Además, esta plataforma también permite compartir los comandos de navegación entre todos los dispositivos. La plataforma en [10] ya utiliza subtítulos en formato Web Video Text Tracks (WebVTT) [11], que es el estándar bajo desarrollo en el W3C (World Wide Web Consortium) para presentar datos multimedia textuales en documentos HTML, y los sincroniza con contenido DASH (Dynamic Adaptive Streaming over HTTP) en directo, utilizando segmentos no multiplexados de la misma duración.

La plataforma presentada en este artículo mejora las plataformas existentes en:

1) Permite la personalización dinámica del formato de los subtítulos y de su posicionamiento, como las plataformas en [6-9], pero beneficiándose del potencial de los componentes HTML5, WebVTT y de las propiedades Cascading Style Sheets (CSS).

2) Proporciona mayor flexibilidad, soporte y rendimiento que la plataforma en [7], ya que utiliza componentes más modernos y estándar para la presentación dinámica y personalizada de subtítulos, como son el elemento *track* (descrito más adelante) y el formato de subtítulos WebVTT. La plataforma en [11] también utiliza WebVTT, pero no proporciona funcionalidades de personalización, posicionamiento dinámico e IDES.

3) La presentación sincronizada de los subtítulos no está ligada a todas las demás animaciones SMIL en el documento HTML (relacionadas o no), como ocurre en [7], sino sólo a la presentación del contenido audiovisual relacionado.

4) Proporciona IDES más precisa que la plataforma en [7], solucionando la ocurrencia de inconsistencias de presentación tras ejecutar comandos de navegación compartidos.

5) No está centrada únicamente para usuarios con deficiencias auditivas, sino también para usuarios con deficiencias visuales, ya que se puede adaptar el formato (tamaño, tipo de letra, color, añadir efectos visuales...) de los subtítulos, tanto en la pantalla principal como en las pantallas secundarias, superando posibles dificultades de legibilidad debido a la distancia con respecto a la pantalla principal.

Además, nuestra plataforma incluye las siguientes novedades con respecto a las demás:

1) Permite modificar la cantidad de líneas de subtítulos que se muestran simultáneamente, así como restablecer la posición del video con un simple *click* sobre una línea específica.

2) Permite aplicar un desfase temporal positivo o negativo a la presentación de subtítulos con respecto al audio asociado.

3) Permite la selección simultánea de dos idiomas en los dispositivos secundarios.

Plataforma de Sincronización y Personalización de Subtítulos

En esta sección se presentan y describen los componentes tecnológicos utilizados para implementar la plataforma, así como las funcionalidades proporcionadas por la misma. Una visión general de la plataforma y de las funcionalidades proporcionadas se puede ver en la Figura 1, mientras que un diagrama de las entidades involucradas y de los mensajes intercambiados se puede ver en la Figura 2.

Tecnologías Web Utilizadas

Esta plataforma se ha implementado mediante el uso exclusivo de tecnologías web estándar, como HTML5 y Javascript. En concreto, se han utilizado cuatro componentes tecnológicos principales para conseguir las funcionalidades buscadas. El primero de ellos es el elemento *video* de HTML5, que permite insertar videos en páginas web, especificando su dirección y su formato (p.ej., códec, resolución...). Además, se utiliza el elemento *track* de HTML5 (hijo del elemento *video*) para presentar, de manera automática, los subtítulos, así como otros datos multimedia en formato texto (p.ej., anotaciones, descripciones, metadatos, capítulos...). El segundo componente utilizado es *Node.js*, un entorno de desarrollo de código abierto, desarrollado en Javascript, para aplicaciones web

cliente-servidor multi-plataforma. *Node.js* proporciona un modelo de comunicación bidireccional basado en eventos para desarrollar aplicaciones distribuidas interactivas y escalables. El tercer componente utilizado es *Socket.IO*, una librería Javascript que permite comunicaciones bidireccionales basadas en eventos entre clientes y un servidor web (*Node.js*). Mediante el uso de *Socket.IO*, se pueden enviar diferentes tipos de mensajes, incluyendo varios tipos de datos, a través de un canal de comunicaciones único. El cuarto componente utilizado es un algoritmo de sincronización de relojes virtual para conseguir una noción coherente del tiempo en el escenario multi-pantalla.

Subtítulos en Formato WebVTT

WebVTT puede proporcionar subtítulos (p.ej., la transcripción o traducción del contenido audiovisual), *captions* o leyendas (p.ej., información textual adicional), así como otra información textual que se deba alinear en el tiempo con el contenido audiovisual, como descripciones, información sobre capítulos u otros metadatos.

Un fichero de subtítulos WebVTT (con extensión *.vtt*) se compone de varios bloques o ítems, llamados *cues* (ver Figura 3). Cada *cue* básicamente contiene un ID (opcional), marcas de tiempo de inicio y final de su presentación (relativas a la posición del video asociado), la información textual y una línea en blanco que indica el final del *cue*. En cada *cue* también se pueden especificar marcas de tiempos (iniciales y finales) para diferentes fragmentos de texto, proporcionando una presentación con estilo karaoke.

WebVTT permite editar el estilo (fuente, tamaño, color...) del texto de cada *cue*, incluso para diferentes fragmentos de texto en cada *cue*, así como configurar la orientación y la dirección de escritura. Además, dispone de mecanismos para asignar diferentes estilos de texto y voces para diferentes oradores, mostrando el nombre de cada uno de ellos.

Mediante las anteriores características, WebVTT proporciona un formato simple y bastante completo (siendo además extensible) para añadir información textual en sistemas multimedia basados en web.

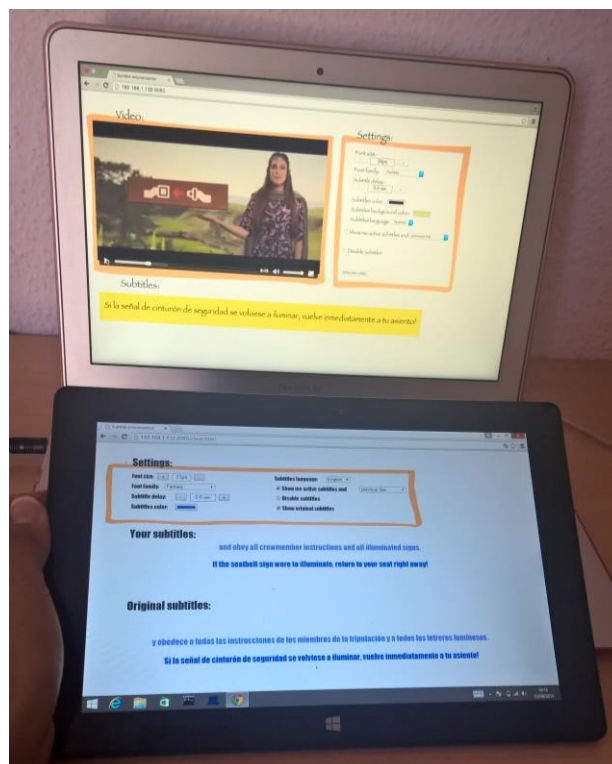


Figura 1. Plataforma de personalización y sincronización de subtítulos en entornos multi-pantalla.

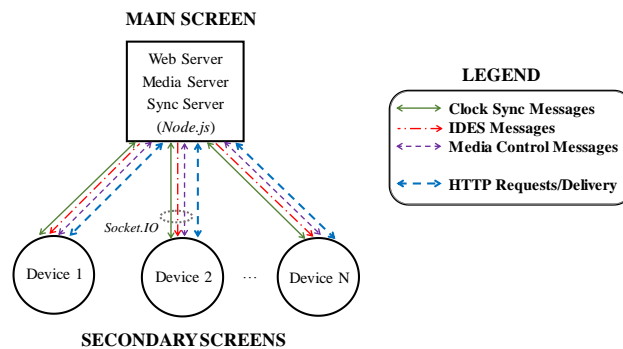


Figura 2. Entidades involucradas y mensajes intercambiados en la plataforma.

Sincronización de Subtítulos con Audio/Video en el mismo dispositivo

Los ficheros de subtítulos WebVTT se enlazan a clips multimedia en un documento HTML (esto es, en la misma instancia del navegador web) a través del elemento o etiqueta estándar *track*, que es hijo del elemento *video* (o *audio*). El elemento *track* tiene varios atributos: i) *src*: la ruta de acceso al fichero WebVTT; ii) *srclang*: el idioma de los subtítulos; iii) una etiqueta visible y legible (p.ej., indicando el idioma); y iv) *kind*: el tipo de track (subtítulos, *captions*, descripciones, capítulos o metadatos).

El elemento *track* será el encargado de presentar cada *cue* de manera sincronizada con el clip multimedia con el que se haya enlazado, de acuerdo a la línea de tiempo local de reproducción de dicho clip. La sincronización entre los *cues* y el clip multimedia vinculado se consigue gracias al evento “*cuechange*” del elemento *track*, sin necesidad de añadir código ni instalar software adicional. Dicha sincronización se mantendrá incluso si se cambia la línea de tiempo de reproducción del clip multimedia a través de comandos de control de navegación, como pausas y saltos a posiciones de video específicas. Un ejemplo de la evolución de la reproducción de un clip de video y un archivo de subtítulos en el mismo dispositivo se puede ver en la gráfica superior de la Figura 3. Se puede observar que la secuencia de video (elemento continuo) se reproduce de manera sincronizada con los *cues* de subtítulos (elementos discretos, cada uno de ellos de una duración y con un intervalo entre *cues* variable). Si se pausa la reproducción, se mantendrá la presentación del bloque de subtítulos (*cue*) correspondiente, mientras que si se salta a una posición específica de video, los *cues* intermedios no se mostrarán.

Sincronización de los Subtítulos en los Dispositivos Secundarios (IDES)

Además de la sincronización entre los subtítulos y el contenido audiovisual en la pantalla principal (o en el mismo dispositivo), nuestra plataforma requiere IDES, es decir, la sincronización entre el contenido multimedia mostrado en la pantalla principal y los subtítulos mostrados en las pantallas secundarias. Se han diseñado dos componentes para conseguirlo: un algoritmo de sincronización de relojes virtual y un protocolo IDES.

Algoritmo de Sincronización de Relojes Virtual

Un requisito para proporcionar IDES es disponer de una noción global y coherente del tiempo en cada uno de las entidades involucradas en el entorno multi-pantalla. Se pueden utilizar tres opciones diferentes para conseguirlo, todas ellas basadas en mecanismos de sincronización de relojes. La primera opción consiste en la sincronización de los relojes del sistema de todas las entidades involucradas (p.ej., mediante el uso de *Network Time Protocol* or NTP). Sin embargo, esta opción podría no ser soportada globalmente en entornos multi-dispositivo y/o multi-plataforma. La segunda opción consiste en la

sincronización de los relojes a nivel de aplicación, pero esto podría implicar la instalación de módulos adicionales (p.ej., clientes NTP Javascript). Además, en ambas opciones, podría suceder que no todas las entidades puedan acceder al mismo servidor de tiempos (p.ej., un servidor NTP específico) o que directamente no soporten la tecnología seleccionada. Por tanto, como tercera opción, se ha diseñado un algoritmo de sincronización de relojes virtual. Básicamente consiste en utilizar un reloj de referencia (p.ej., el del servidor *Node.js*), e ir midiendo los retardos y las desviaciones entre los relojes de las entidades involucradas mediante el envío de mensajes bidireccionales periódicos, vía el canal *Socket.IO* (véase la Figura 2), de una manera similar a como lo hace el protocolo NTP. Con esto se consigue alinear las referencias temporales proporcionadas por los relojes de las entidades involucradas (aplicando los ajustes necesarios para compensar las desviaciones medidas), incluso cuando no utilizan la misma tecnología para la obtención de tiempos globales.

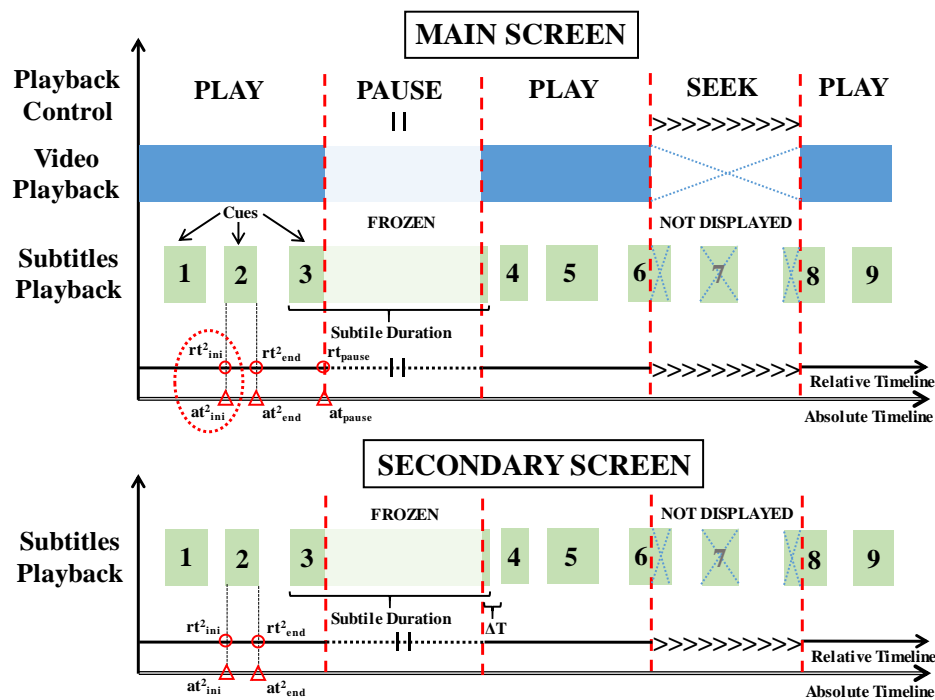


Figura 3. Ejemplo de la reproducción sincronizada entre subtítulos y video.

Protocolo IDES

Además del algoritmo de sincronización de relojes virtual, se ha diseñado e implementado un protocolo IDES, que adopta un esquema de control Master/Slave (M/S), en el que el dispositivo principal (en el que se aloja el servidor *Node.js*) es el Master y los dispositivos secundarios son los Slaves. Otros esquemas centralizados o distribuidos podrían haberse implementado, pero el esquema M/S es el más simple y apropiado para el tipo de escenarios multi-pantalla bajo estudio [12]. En este esquema, el dispositivo Master enviará de manera periódica (unicast o multicast), vía el canal *Socket.IO* (véase Figura 2), mensajes de control a los dispositivos Slaves. El periodo de envío de dichos mensajes se puede configurar en la plataforma (por defecto, será de 1s) y se adaptará de manera dinámica en función del número de dispositivos secundarios activos, con tal de no saturar los recursos de red ni los recursos computacionales en el dispositivo Master. Estos mensajes IDES también se enviarán cada vez que ocurre un evento "cuechange", que significa que una nueva *cue* debe presentarse. Dichos mensajes IDES contienen dos campos con marcas de tiempo para conseguir la sincronización. El primer campo, *rt* (*relative timestamp*), indica la posición de reproducción del video (y de los subtítulos) en la pantalla principal, en base a la línea de tiempos relativa (local) del reproductor. El segundo campo, *at* (*absolute timestamp*), es el tiempo absoluto (global) asociado al momento en el que se

obtuvo rt . La inclusión y asociación entra ambas marcas de tiempo (véase Figura 3), junto con el algoritmo de sincronización de relojes virtual, permite compensar el retardo de tránsito de los mensajes IDES cuando se reciben en cada uno de los dispositivos secundarios (Slaves). En ese momento, cada dispositivo Slave calculará la asincronía (es decir, la diferencia temporal) entre su proceso de reproducción y el del dispositivo Master. Si dicha asincronía supera un umbral pre-establecido (configurable en la plataforma), el dispositivo Slave debe ajustar su proceso de reproducción para eliminarla. Esto se puede conseguir mediante dos tipos de técnicas de ajuste del proceso de reproducción. La primera de ellas consiste en ejecutar saltos y pausas de una magnitud igual a la asincronía detectada. La segunda de ellas consiste en ajustar de manera suavizada la tasa de reproducción (es decir, acelerando o ralentizando dentro de límites permisibles) durante un intervalo de tiempo necesario para eliminar la asincronía detectada. Esta segunda técnica es mucho más conveniente ya que proporciona un ajuste de sincronización más preciso y permite evitar la ocurrencia de cambios bruscos en la evolución del proceso de reproducción, que pueden ser molestos para los usuarios (mala QoE).

Adicionalmente, nuestra plataforma también permite compartir los comandos de control de navegación entre todos los dispositivos, añadiendo también marcas de tiempo (relativas y absolutas) de envío en dichos mensajes para conseguir mejor precisión de sincronización. Con esto se consiguen sesiones interactivas. Por ejemplo, el video y los subtítulos se pueden pausar en todos los dispositivos para comentar una escena específica o comprobar cómo se escriben ciertas palabras, o bien se puede volver a visualizar una escena para volver a comprobar cómo se escriben o pronuncian palabras o frases específicas. Gracias a la bidireccionalidad del canal *Socket.IO* y a su comportamiento basado en eventos, también es posible permitir la ejecución de los comandos de control de navegación en los dispositivos secundarios, si se especifica en el menú de ajustes de la plataforma. En este caso, dichos comandos serán re-enviados por el dispositivo principal (servidor *Node.js*) a todos los dispositivos secundarios involucrados.

Aparte de proporcionar sincronización precisa, la inclusión de ambos tipos de marcas de tiempo en dichos comandos de control de navegación permite evitar la presentación inconsistente de subtítulos en los dispositivos involucrados después de ejecutarlos, como ocurre en la plataforma en [7]. Por ejemplo, en esa plataforma, tal y como reconocen sus autores, cuando se pausa el vídeo en la pantalla principal, si el retardo de red no es muy pequeño, es posible que se muestre la siguiente *cue* en los dispositivos secundarios durante la pausa, especialmente cuando este comando se ejecuta justo al final de un *cue*. Por ejemplo, en la gráfica inferior de la Figura 3, si el retardo es superior a ΔT , la cuarta *cue* se mostraría durante la pausa, en vez de la tercera. Estas situaciones indeseadas se evitan en nuestra plataforma gracias a la inclusión de las marcas de tiempo en los mensajes.

Finalmente, los usuarios también pueden ajustar un desfase temporal (positivo o negativo) a la presentación de los subtítulos con respecto a la presentación del audio/vídeo. Esto puede ser útil en algunas situaciones. Por ejemplo, los usuarios podrían preferir leer los subtítulos un poco antes o después que el audio asociado para comprobar cómo se pronuncian las palabras (o frases) o si son capaces de identificarlas, respectivamente.

Personalización del Formato y Posicionamiento de los Subtítulos

La personalización del estilo (p.ej., formato, tamaño, posición, orientación...) de cada *cue* se puede conseguir con las propiedades de WebVTT (la mayoría de ellas comentadas previamente). Además, se ha adoptado otra estrategia para proporcionar mayor flexibilidad, más posibilidades y mayor control dinámico para la personalización de los subtítulos, especialmente cuando están almacenados en ficheros remotos, y/o generados por terceras partes. Como se ha mencionado previamente, el elemento *track* puede proporcionar diferentes tipos de datos multimedia textuales, especificados mediante su

atributo *kind*. Si se asigna a este atributo la opción “*metadata*”, la información de los *cues* no es visible para los usuarios, pero se puede leer, interpretar y presentar en otros elementos HTML, como en un *div*. Mediante esta estrategia, nuestra plataforma posibilita la personalización dinámica del formato de los subtítulos (p.ej., color, fuente, tamaño...), utilizando la amplia gama de propiedades CSS, así como su orientación y posición (p.ej., arrastrar y soltar, posicionamiento dinámico, incluso más allá de la ventana del reproductor o de la pantalla principal...) con tal de optimizar su legibilidad, el confort de los usuarios y/o la estética de la aplicación. Además, esta estrategia también permite la adaptación del formato y posicionamiento de los subtítulos en base a la evolución dinámica de la aplicación o de las condiciones del entorno. Por ejemplo, la región de subtítulos se puede re-dimensionar automáticamente si el tamaño de la ventana del reproductor cambia.

La personalización dinámica de los subtítulos no afectará a su sincronización con el contenido audiovisual.

Selección del Idioma de los Subtítulos

Nuestra plataforma también permite seleccionar el idioma de los subtítulos, a través de una lista desplegable que indica los idiomas disponibles (véase la Figura 4), tanto en la pantalla principal como en las secundarias. La selección del idioma asignará el valor correspondiente a los atributo *src* and *srclang* del elemento *track*. Esta funcionalidad también se proporciona en la plataforma en [7], pero la nuestra aprovecha la existencia de los atributos anteriores, lo que ofrece una solución más ligera, flexible y rápida que tener que solicitar y cargar un fichero de subtítulos SRT remoto y transformarlo en un componente SVG, como se hace en [7]. Cuando el nuevo fichero de subtítulos se haya cargado, los subtítulos en el idioma seleccionado se presentarán de manera sincronizada con el contenido audiovisual adicional, tanto en el mismo como en otros dispositivos.

Además, en los dispositivos secundarios, a pesar de que sólo se dispone de un elemento *track* por cada elemento *video*, cada usuario también puede habilitar la presentación de los subtítulos mostrados en la pantalla principal, que serán enviados a través del canal *Socket.IO* y mostrados en otro elemento *div* de manera sincronizada (véase la Figura 1).

Navegación de Subtítulos

Tanto en la pantalla principal como en la secundaria (y de manera independiente), se puede configurar el número de líneas de subtítulos (*n*) que se deben mostrar simultáneamente. Esto se consigue cargando las *cues* apropiadas (p.ej. las *n*-1 previas a la *cue* active). En este caso, los usuarios también pueden hacer click sobre una línea de subtítulos específica para saltar a la posición del vídeo correspondiente. Esto es otra novedad de nuestra plataforma, que puede ser muy útil en escenarios de aprendizaje de idiomas, o para volver a ver o escuchar una escena específica en caso de distracción.

Un video mostrando las funcionalidades de la plataforma está disponible en <https://goo.gl/xS9HVX>.

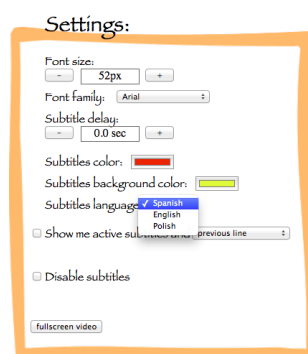


Figura 4. Selección del idioma de los subtítulos.

Trabajo Futuro

Varias extensiones y mejoras hay planificadas para futuras versiones de la plataforma. En primer lugar, queremos explorar los desafíos y las soluciones potenciales para posibilitar la generación y presentación personalizada y dinámica de subtítulos para contenidos en vivo. En segundo lugar, queremos añadir funcionalidades de inteligencia artificial, como síntesis y reconocimiento de voz. Esto permitirá la generación automática tanto de audio como de subtítulos para idiomas adicionales a los disponibles en los archivos multimedia. En tercer lugar, queremos adaptar la plataforma para que sea compatible con el estándar HbbTV. En cuarto lugar, queremos evaluar de manera objetiva y subjetiva el rendimiento y usabilidad de la plataforma en diferentes escenarios.

Agradecimientos

El presente trabajo ha sido financiado, en parte, por el FEDER y por el MINECO, bajo el programa de apoyo a la I+D+i en el proyecto con referencia TEC2013-45492-R.

Referencias

- [1] M. Armstrong, "The Development of a Methodology to Evaluate the Perceived Quality of Live TV Subtitles", BBC Research & Development, White Paper WHP 259, Sept. 2013.
- [2] Pereira, A., "Criteria for Elaborating Subtitles for Deaf and Hard of Hearing Adults in Spain: Description of a Case Study". In *Listening to Subtitles. Subtitles for the Deaf and Hard of Hearing*; Matamala, A., Orero, P., Eds.; Ed. Peter Lang; ISBN 978-3-0343-0353-8, pp. 87–102, 2010.
- [3] Lorenzo, L., "Criteria for Elaborating Subtitles for Deaf and Hard of Hearing Children in Spain: A Guide of Good Practice". In *Listening to Subtitles. Subtitles for the Deaf and Hard of Hearing*; Matamala, A., Orero, P., Eds.; Ed. Peter Lang; Bern, ISBN 978-3-0343-0353-8, pp. 139–148, 2010.
- [4] Porteiro, M., "The Use of Subtitles to Treat Speech-Language Disorders", *Perspectives: Studies in Translatology*, Vol. 21, No. 1, 1-12, 2012.
- [5] Lavour, J.M., Bairstow, D., "Languages on the Screen: Is Film Comprehension Related to the Viewers' Fluency Level and to the Language in the Subtitles?", *Int. J. Psychol*, 46, pp. 455–462, 2011.
- [6] Hong, R., et al., "Video accessibility enhancement for hearing-impaired users", *ACM TOMCCAP*, 7S, 1, Article 24, 19 pages, November 2011.
- [7] Rodriguez, A., Talavera, G., Orero, P., Carrabina, J., "Subtitle Synchronization across Multiple Screens and Devices", *Sensors*, 12(7), 8710-8731, June 2012.
- [8] Y. Hu, J. Kautz, Y. Yu, and W. Wang. "Speaker-Following Video Subtitles", *ACM TOMCCAP*, 11, 2, Article 32, 17 pages, January 2015.
- [9] Brown, A., Jones, R., Crabb, M., "Dynamic Subtitles: the User Experience", *ACM TVX 2015*, Brussels (Belgium), June 2015
- [10] Concolato, C., Le Feuvre, J., *Live HTTP streaming of video and subtitles within a browser*, MMSys '13, 146-150, Oslo (Norway), February 2013.
- [11] WebVTT: The Web Video Text Tracks Format, W3C Community Group Draft, S. Pfeiffer, P. Jägenstedt, I. Hickson (Editors), <http://dev.w3.org/html5/webvtt/>, July 2015.
- [12] Montagud, M., et al. 2012. Inter-destination multimedia synchronization: schemes, use cases and standardization. *MMSJ*, 18(6), 459-482, November 2012.